

RAW SEQUENCE LISTING

**The Biotechnology Systems Branch of the Scientific and Technical
Information Center (STIC) no errors detected.**

Application Serial Number: 10/540,405
Source: JFCW
Date Processed by STIC: 02/21/2007

ENTERED



IFWO

RAW SEQUENCE LISTING

DATE: 02/21/2007

PATENT APPLICATION: US/10/540,405

TIME: 14:18:33

Input Set : E:\9301210999.txt

Output Set: N:\CRF4\02212007\J540405.raw

3 <110> APPLICANT: Rosetta Inpharmatics LLC
 4 Schadt, Eric
 5 Monks, Stephanie
 7 <120> TITLE OF INVENTION: COMPUTER SYSTEMS AND METHODS FOR ASSOCIATING GENES WITH
 TRAITS
 8 USING CROSS SPECIES DATA
 10 <130> FILE REFERENCE: 9301-210-228
 C--> 12 <140> CURRENT APPLICATION NUMBER: US/10/540,405
 C--> 13 <141> CURRENT FILING DATE: 2005-06-22
 15 <150> PRIOR APPLICATION NUMBER: 60/436,684
 16 <151> PRIOR FILING DATE: 2002-12-27
 18 <150> PRIOR APPLICATION NUMBER: 60/460,343
 19 <151> PRIOR FILING DATE: 2003-04-03
 21 <160> NUMBER OF SEQ ID NOS: 30
 23 <170> SOFTWARE: PatentIn version 3.2
 25 <210> SEQ ID NO: 1
 26 <211> LENGTH: 1583
 27 <212> TYPE: DNA
 28 <213> ORGANISM: Mus musculus
 30 <400> SEQUENCE: 1

31	gagctattcg	gcctctctag	gccggcggggt	cctccgctcc	atggctctgt	ctgtcagcgc	60
33	tgtgtcagga	ggccagtgcc	gaggtccggg	cgcgctccga	cgcttcgacc	ctcgagccgg	120
35	tcgcgggat	cccggcgggc	gcgggacgat	ggcgtgggtg	cactgacagg	cgcgggcggc	180
37	tgccgagccc	cgcgggcggc	atggcgggcc	agttccgcag	ctacgtgtgg	gacccgttgc	240
39	taatcctgtc	gcagatcgta	ctcatgcaga	ccgtctacta	tggctctctg	ggcctgtggc	300
41	tggcgctggg	ggacgcgctg	gtgcgcaagc	ccgtccctgg	accagatgtt	cgacgcggag	360
43	atcctgggct	tctccacccc	tccaggccgg	ctctcaatga	tgtccttcgt	cctcaacgcc	420
45	ctcacctgtg	ccctgggctt	gctgtacttc	atccggcgag	ggaagcagtg	cctggatttc	480
47	actgtcactg	tgcatttctt	tcacctcctg	ggctgctggc	tctacagctc	ccgtttcccc	540
49	tcggcgctga	cctgggtggc	ggtccaggct	gtgtgcattg	cactcatggc	cgatcatcgg	600
51	gagtacctgt	gcattgcggc	ggagctcaag	gagatcccc	tcagctcagc	ccctaagtcc	660
53	aatgtctaga	gttggggcct	ttggacactc	tgctggcact	tggggcccat	caccttgggc	720
55	tgctcagacc	tccagatggg	gtctggccca	agtctgagca	gaaccttgga	aatgtgaagt	780
57	ctgttggtgg	agagataatg	aggtcccac	ataaaggcag	gtagcagcca	tgatcacaga	840
59	tgtaagaatg	gcctctgtct	gccaaagcct	tgatatctgg	aggccagtaa	gggacctcat	900
61	ggagggtagt	ggcagatttg	gaaccatgtc	acatgagcca	tcatactgtc	accagcctgt	960
63	tattttaaaa	agaaaaaaa	aaaatcaagg	atatctgatt	ggaataaacc	actcttctcg	1020
65	ttgtctgtct	tatgcccatt	acagccagta	cctttgctgt	gttgccaaac	cacagggatt	1080
67	ctctgtggag	aaataacctg	tttctgggtc	catagccaca	gaaaaagatg	taggtacaga	1140
69	gtgctaggct	gctgacagga	cgctcgaggg	aggaggcatc	aagcacaaga	aaaatgcatg	1200
71	gccgtgccgt	tagacacaca	cacacacttt	tgtgtgtgtc	caggacccat	gactgtctcc	1260
73	ctccagttcc	ctgtatggac	tctgccttgc	tgttgtcact	cagcacagcc	agagacagga	1320
75	cccagagaaa	accccagcat	ccctcccagc	cttcccttca	taataaaagc	cattgtctgc	1380
77	tctctggaag	tgagcaggca	gccagcttct	actggacctc	aactgtggca	ggagtttctg	1440

(P8-6)

RAW SEQUENCE LISTING

DATE: 02/21/2007

PATENT APPLICATION: US/10/540,405

TIME: 14:18:33

Input Set : E:\9301210999.txt

Output Set: N:\CRF4\02212007\J540405.raw

```

79 tttgctgtct tttgagttct gtgatagga ggggtgtacta aagggtgctgg aggctcaccc 1500
81 tgctaagctt tcttccaagt ggtttcctca ggaagggtcg gcagctgtcc ttcctaggta 1560
83 cataaatata ctattttcca atc 1583
86 <210> SEQ ID NO: 2
87 <211> LENGTH: 2677
88 <212> TYPE: DNA
89 <213> ORGANISM: Homo sapiens
91 <400> SEQUENCE: 2
92 tctaggccgg cagcgctct cctccatggc cctgtctgtc agcgctgttt tgggagcccg 60
94 ccggtgaggg cgggccacgc tcagacactt cgatcgatga gtctgtcact gggcatggcg 120
96 ggtcagttcc gcagctacgt gtgggacccg ctgctgatcc tgtcgcagat cgtcctcatg 180
98 cagaccgtgt attacggctc gctgggcctg tggctggcgc tgggtggacgg gctagtgcga 240
100 cagcccctcg ctggaccaga tgttcgacgc cgagatcctg ggcttttcca ccctccagg 300
102 ccggtctctc atgatgtcct tcactcctca cggcctcacc tgtgccctgg gcttgcctga 360
104 cttcatccgg cgaggaaagc agtgtctgga tttcactgtc actgtccatt tctttcacct 420
106 cctgggctgc tggttctaca gctcccgttt ccctcggcg ctgacctggg ggctgggtcca 480
108 agccgtgtgc attgcactca tggctgtcat cggggagtag ctgtgcatgc ggacggagct 540
110 caaggagata ccctcaact cagcccctaa atccaatgtc tagaatcagg ccctttggac 600
112 atcctgctga cacttggggc ccttaacacc ttgggctgct cagacgctcc agatgagggtc 660
114 cagcccagat ctgagaggaa ccctggaaat gtgaagtctc tgttggtttg ggagagatag 720
116 tgagggcctg tcaaagaagg caggtagcag tcagcatgac agctgcaaga atgacctctg 780
118 tctgttgaa g ccttggatc tgagagggtca ggaaggggac ctctttgagg gtaataacag 840
120 aattggaacc atgccactct tgagccacaa tacctgtcac cagcctgttg ttttaagaga 900
122 gaaaaaaaaat caaggatatc tgattggagc aaaccacttc tttagtcatc tgtcttacct 960
124 ccctgggaca gctgttacct ttgcagtgtt gccgaatcac agcagttacc tttgcagtgt 1020
126 tgccgaatca cagcagttct gttggagaaa cgcttggttt ccggatccag agccacagaa 1080
128 agaaatgtag gtgtgaagta ttaggtgtgt gtcagggaga ggatggcaga tggaggcatc 1140
130 aagcacaagg aaaatgcaca acctgtgccc tgttatcac acgttcattg gcaccaaga 1200
132 acctatgact ttcttccagt tccttctacc aggtcccat cctgctgcca gctctcaaca 1260
134 tagcaggcca taggaccagc agaagaatcc cagcgttgct caaagtctaa ccatacataa 1320
136 gacactgcct gtcttctagg aatgaccagg caccagctc ccactggact ccaatttttt 1380
138 ttctgcctt atttagaatt ctttggcggg aagggtatga tgggttccca gagacaagaa 1440
140 gccaacctt ctggcctggg ctgtgctgat agtgcgtagg gagataggaa tttgctgcta 1500
142 agatttttct ttgggggtga gtttcctctg tgaggggctt gcagctatcc ttcctgtgta 1560
144 taaaaataca gtattttcca tggttctgcc tgcacttact ttgtaatgcc acggttgaga 1620
146 ttgagagaga tcagcgcagc caggcaaggg aactttaaa aattattagg ccaccttctc 1680
148 ctttctctgg accccagagt cattcctcca tttggttaaa atactcagtg cagggaaactc 1740
150 ttacatcctg tctccttcac ttgcagcgtc ccctgctatg cctcagggtga accacataat 1800
152 tcttgggttt ccgttcctac ttgctagtga tttctgaaca tgttcaatgg agcggcacac 1860
154 agtctagacc cacttccgca ttgaaacctt cactgttcct ctttggtttc ttcagagctt 1920
156 tcccaagaga gctgtcagtt ttcagctgtc agtaacacaa atgagtttat ggtaacacaa 1980
158 atgagttttg ctatctctct gagaagctca tctgacctcc tgactctcag ccctacagag 2040
160 tagggagttg atgctgacag gatgaagatt taggaataaa tatgcctggg aagagactgg 2100
162 gaaggttcta gggtgaggca cctcagtaac tcatgggtacc ttggccaagt tggaaagaa 2160
164 cagtttgta atgaggcaca gtaatcctgg ctgcagggtc taggaggtaa gaccagctgg 2220
166 gatgaccttc cctgggttaa tcaatttccc tctagacaac acaaaactgca ggcatgtgac 2280
168 taactttgaa agaacaacca tcatgtggct gctgtcaccc ttgaccagcc gtggtggtgg 2340
170 ttactccatc tgtggttgga gcgcctctt gggattcact tcaaggctct gtgcctattt 2400
172 ttctgcatat cttctgtgat gacaaatctc tgtccctga gtgttaattt gatttttaga 2460

```

RAW SEQUENCE LISTING

DATE: 02/21/2007

PATENT APPLICATION: US/10/540,405

TIME: 14:18:33

Input Set : E:\9301210999.txt

Output Set: N:\CRF4\02212007\J540405.raw

```

174 aatggccaaa agtcacgtga tccaaacttt ttttcagtaa tatggagact gagctgcatg 2520
176 gtagttgggg atcaaaaata tgtgacctta atgagatttt tatgatttct aaagtaacaa 2580
178 taaaagcagt ttttagagtt gagttccaga gagggcaggg caatggcagt gacatgtttg 2640
180 tcattttaat aataaataac atctattgag tgcttaa 2677

```

183 <210> SEQ ID NO: 3

184 <211> LENGTH: 453

185 <212> TYPE: DNA

186 <213> ORGANISM: Homo sapiens

188 <400> SEQUENCE: 3

```

189 atggcggtc agttccgcag ctacgtgtgg gacccgctgc tgatcctgtc gcagatcgtc 60
191 ctcatgcaga ccgtgtatta cggctcgctg ggctgtggc tggcgctggt ggacgggcta 120
193 gtgcgacagc ccctcgctgg accagatgtt cgacgccgag atcctgggct tttccacccc 180
195 tccaggccgg ctctccatga tgtccttcac cctcaacgcc ctacactgtg cctggggctt 240
197 gctgtacttc atccggcgag gaaagcagtg tctggatttc actgtcactg tccatttctt 300
199 tcacctcctg ggctgctggt tctacagctc ccgtttcccc tcggcgctga cctgggtggct 360
201 ggtccaagcc gtgtgcattg cactcatggc tgtcatcggg gagtacctgt gcatgcggac 420
203 ggagctcaag gagatacccc tcaactcagc ccc 453

```

206 <210> SEQ ID NO: 4

207 <211> LENGTH: 156

208 <212> TYPE: PRT

209 <213> ORGANISM: Mus musculus

211 <400> SEQUENCE: 4

```

213 Met Ala Gly Gln Phe Arg Ser Tyr Val Trp Asp Pro Leu Leu Ile Leu
214 1 5 10 15
217 Ser Gln Ile Val Leu Met Gln Thr Val Tyr Tyr Gly Ser Leu Gly Leu
218 20 25 30
221 Trp Leu Ala Leu Val Asp Ala Leu Val Arg Ser Ser Pro Ser Leu Asp
222 35 40 45
225 Gln Met Phe Asp Ala Glu Ile Leu Gly Phe Ser Thr Pro Pro Gly Arg
226 50 55 60
229 Leu Ser Met Met Ser Phe Val Leu Asn Ala Leu Thr Cys Ala Leu Gly
230 65 70 75 80
233 Leu Leu Tyr Phe Ile Arg Arg Gly Lys Gln Cys Leu Asp Phe Thr Val
234 85 90 95
237 Thr Val His Phe Phe His Leu Leu Gly Cys Trp Leu Tyr Ser Ser Arg
238 100 105 110
241 Phe Pro Ser Ala Leu Thr Trp Trp Leu Val Gln Ala Val Cys Ile Ala
242 115 120 125
245 Leu Met Ala Val Ile Gly Glu Tyr Leu Cys Met Arg Thr Glu Leu Lys
246 130 135 140
249 Glu Ile Pro Leu Ser Ser Ala Pro Lys Ser Asn Val
250 145 150 155

```

253 <210> SEQ ID NO: 5

254 <211> LENGTH: 125

255 <212> TYPE: PRT

256 <213> ORGANISM: Mus musculus

258 <400> SEQUENCE: 5

```

260 Met Ala Leu Trp Ala Cys Gly Trp Arg Trp Trp Thr Arg Trp Cys Ala
261 1 5 10 15

```

RAW SEQUENCE LISTING

DATE: 02/21/2007

PATENT APPLICATION: US/10/540,405

TIME: 14:18:33

Input Set : E:\9301210999.txt

Output Set: N:\CRF4\02212007\J540405.raw

```

264 Gln Pro Val Pro Gly Pro Asp Val Arg Arg Gly Asp Pro Gly Leu Leu
265          20          25          30
268 His Pro Ser Arg Pro Ala Leu Asn Asp Val Leu Arg Pro Gln Arg Pro
269          35          40          45
272 His Leu Cys Pro Gly Leu Ala Val Leu His Pro Ala Arg Glu Ala Val
273          50          55          60
276 Pro Gly Phe His Cys His Cys Ala Phe Leu Ser Pro Pro Gly Leu Leu
277 65          70          75          80
280 Ala Leu Gln Leu Pro Phe Pro Leu Gly Ala Asp Leu Val Ala Gly Pro
281          85          90          95
284 Gly Cys Val His Cys Thr His Gly Arg His Arg Gly Val Pro Val His
285          100          105          110
288 Ala Asp Gly Ala Gln Gly Asp Pro Pro Gln Leu Ser Pro
289          115          120          125
292 <210> SEQ ID NO: 6
293 <211> LENGTH: 151
294 <212> TYPE: PRT
295 <213> ORGANISM: Mus musculus
297 <400> SEQUENCE: 6
299 Met Ala Gly Gln Phe Arg Ser Tyr Val Trp Asp Pro Leu Leu Ile Leu
300 1          5          10          15
303 Ser Gln Ile Val Leu Met Gln Thr Val Tyr Tyr Gly Ser Leu Gly Leu
304          20          25          30
307 Trp Leu Ala Leu Val Asp Ala Leu Val Arg Lys Pro Val Pro Gly Pro
308          35          40          45
311 Asp Val Arg Arg Gly Asp Pro Gly Leu Leu His Pro Ser Arg Pro Ala
312          50          55          60
315 Leu Asn Asp Val Leu Arg Pro Gln Arg Pro His Leu Cys Pro Gly Leu
316 65          70          75          80
319 Ala Val Leu His Pro Ala Arg Glu Ala Val Pro Gly Phe His Cys His
320          85          90          95
323 Cys Ala Phe Leu Ser Pro Pro Gly Leu Leu Ala Leu Gln Leu Pro Phe
324          100          105          110
327 Pro Leu Gly Ala Asp Leu Val Ala Gly Pro Gly Cys Val His Cys Thr
328          115          120          125
331 His Gly Arg His Arg Gly Val Pro Val His Ala Asp Gly Ala Gln Gly
332          130          135          140
335 Asp Pro Pro Gln Leu Ser Pro
336 145          150
339 <210> SEQ ID NO: 7
340 <211> LENGTH: 156
341 <212> TYPE: PRT
342 <213> ORGANISM: Mus musculus
344 <400> SEQUENCE: 7
346 Met Ala Gly Gln Phe Arg Ser Tyr Val Trp Asp Pro Leu Leu Ile Leu
347 1          5          10          15
350 Ser Gln Ile Val Leu Met Gln Thr Val Tyr Tyr Gly Ser Leu Gly Leu
351          20          25          30
354 Trp Leu Ala Leu Val Asp Ala Leu Val Arg Ser Ser Pro Ser Leu Asp

```

RAW SEQUENCE LISTING

DATE: 02/21/2007

PATENT APPLICATION: US/10/540,405

TIME: 14:18:33

Input Set : E:\9301210999.txt

Output Set: N:\CRF4\02212007\J540405.raw

```

355          35          40          45
358 Gln Met Phe Asp Ala Glu Ile Leu Gly Phe Ser Thr Pro Pro Gly Arg
359          50          55          60
362 Leu Ser Met Met Ser Phe Val Leu Asn Ala Leu Thr Cys Ala Leu Gly
363 65          70          75          80
366 Leu Leu Tyr Phe Ile Arg Arg Gly Lys Gln Cys Leu Asp Phe Thr Val
367          85          90          95
370 Thr Val His Phe Phe His Leu Leu Gly Cys Trp Leu Tyr Ser Ser Arg
371          100          105          110
374 Phe Pro Ser Ala Leu Thr Trp Trp Leu Val Gln Ala Val Cys Ile Ala
375          115          120          125
378 Leu Met Ala Val Ile Gly Glu Tyr Leu Cys Met Arg Thr Glu Leu Lys
379          130          135          140
382 Glu Val Pro Leu Ser Ser Ala Pro Lys Ser Asn Val
383 145          150          155
386 <210> SEQ ID NO: 8
387 <211> LENGTH: 151
388 <212> TYPE: PRT
389 <213> ORGANISM: Homo sapiens
391 <400> SEQUENCE: 8
393 Met Ala Gly Gln Phe Arg Ser Tyr Val Trp Asp Pro Leu Leu Ile Leu
394 1          5          10          15
397 Ser Gln Ile Val Leu Met Gln Thr Val Tyr Tyr Gly Ser Leu Gly Leu
398          20          25          30
401 Trp Leu Ala Leu Val Asp Gly Leu Val Arg Gln Pro Leu Ala Gly Asp
402          35          40          45
405 Pro Val Arg Arg Arg Asp Pro Gly Leu Phe His Pro Ser Arg Pro Ala
406          50          55          60
409 Leu His Asp Val Leu His Pro Gln Arg Pro His Leu Cys Pro Gly Leu
410 65          70          75          80
413 Ala Val Leu His Pro Ala Arg Lys Ala Val Ser Gly Phe His Cys His
414          85          90          95
417 Cys Pro Phe Leu Ser Pro Pro Gly Leu Leu Val Leu Gln Leu Pro Phe
418          100          105          110
421 Pro Leu Gly Ala Asp Leu Val Ala Gly Pro Ser Arg Val His Cys Thr
422          115          120          125
425 His Gly Cys His Arg Gly Val Pro Val His Ala Asp Gly Ala Gln Gly
426          130          135          140
429 Asp Thr Pro Gln Leu Ser Pro
430 145          150
433 <210> SEQ ID NO: 9
434 <211> LENGTH: 3685
435 <212> TYPE: DNA
436 <213> ORGANISM: Mus musculus
439 <220> FEATURE:
440 <221> NAME/KEY: CDS
441 <222> LOCATION: (52)..(3195)
443 <220> FEATURE:
444 <221> NAME/KEY: misc_feature

```

RAW SEQUENCE LISTING ERROR SUMMARY
PATENT APPLICATION: US/10/540,405

DATE: 02/21/2007
TIME: 14:18:34

Input Set : E:\9301210999.txt
Output Set: N:\CRF4\02212007\J540405.raw

Please Note:

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:9; N Pos. 3406

VERIFICATION SUMMARY

DATE: 02/21/2007

PATENT APPLICATION: US/10/540,405

TIME: 14:18:34

Input Set : E:\9301210999.txt

Output Set: N:\CRF4\02212007\J540405.raw

L:12 M:270 C: Current Application Number differs, Replaced Current Application Number

L:13 M:271 C: Current Filing Date differs, Replaced Current Filing Date

L:723 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:9 after pos.:3395